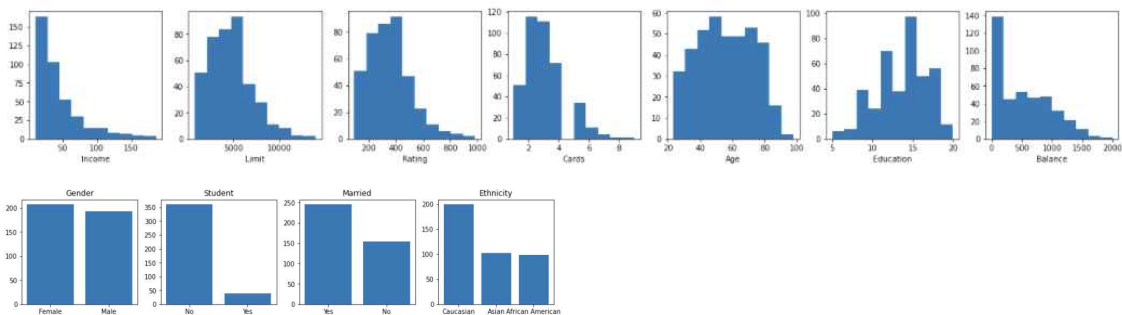


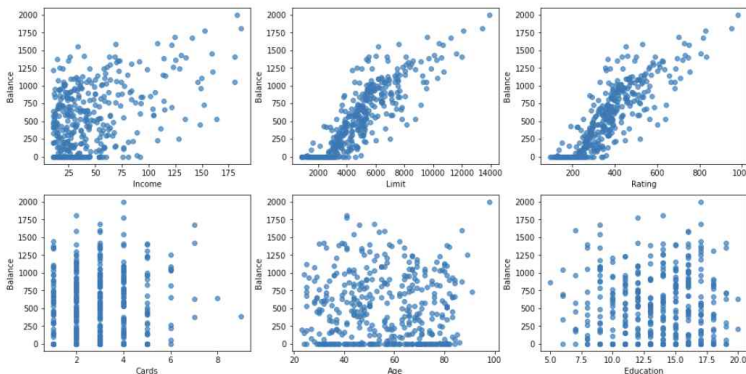
## &lt; 과제 #3 : Regression &gt;

HW3.csv 파일에는 은행고객의 수입(Income), 신용한도(Limit), 신용점수(Rating), 신용카드 개수(Cards), 나이(Age), 교육년수(Education), 성별(Gender), 학생 여부(Student), 결혼 여부(Married), 인종(Ethnicity), 그리고 잔고(Balance)가 저장되어 있다. Balance를 레이블로 하는 아래 회귀분석을 실시하시오.

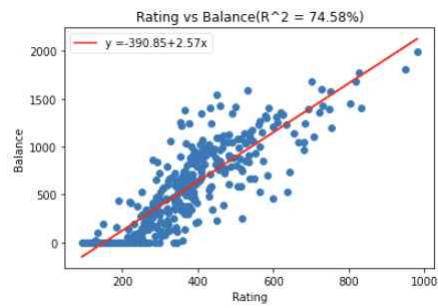
- 1) 파일을 읽어 데이터프레임을 생성하고 처음 3개의 샘플을 출력하시오. 특성의 개수와 샘플 (관측치)의 개수를 구하고 결측치가 있는지 알아보시오. 수치형 특성과 범주형 특성은 무엇 인지 파악하시오.
- 2) 수치형 특성과 레이블에 대하여는 히스토그램을 작성하고 범주형 특성에 대하여는 막대그 래프를 작성하고 간단히 코멘트 하시오.



- 3) 데이터프레임에 `corr()` 메서드를 적용하여 변수들간의 상관계수를 출력하고 해석하시오.
- 4) 각 수치형 특성과 Balance(잔고) 변수와의 산점도를 아래와 같은 subplot으로 작성하시오.



- 5) Rating(신용점수) 변수를 특성으로 하고 Balance(잔고) 변수를 레이블로 하는 단순선형회귀모형을 훈련하고 산점도에 회귀직선식을 표시하고 이 모형의 결정계수( $R^2$ )을 구하시오.



- 6) 수치형 특성들만 포함하여 데이터를 훈련용과 테스트용으로 7:3으로 분할 (random\_state=0)하여 선형회귀모형을 훈련시키고 훈련데이터와 테스트데이터에 대한 결정계수를 구하시오. 과대적합이나 과소적합이 의심되는지 코멘트 하시오. Income이 100, Limit가 6000, Rating이 550, Card가 5개, 나이가 55, 교육년수가 20인 사람의 잔고를 예측하시오.